

1647 #13  
5311/20/02  
1600

## RAW SEQUENCE LISTING

DATE: 10/18/2002

PATENT APPLICATION: US/09/508,997A

TIME: 10:41:41

Input Set : D:\seqlist.txt

Output Set: N:\CRF4\10182002\I508997A.raw

RECEIVED

OCT 25 2002

TECH CENTER 1600/2900

ENTERED

3 <110> APPLICANT: MIYATA, Toshio  
 5 <120> TITLE OF INVENTION: Megsin Protein  
 7 <130> FILE REFERENCE: SHIM-004  
 9 <140> CURRENT APPLICATION NUMBER: 09/508,997A  
 10 <141> CURRENT FILING DATE: 2000-06-27  
 12 <150> PRIOR APPLICATION NUMBER: JP 09-275302  
 13 <151> PRIOR FILING DATE: 1997-09-22  
 15 <160> NUMBER OF SEQ ID NOS: 44  
 17 <170> SOFTWARE: PatentIn Ver. 2.0  
 19 <210> SEQ ID NO: 1  
 20 <211> LENGTH: 1143  
 21 <212> TYPE: DNA  
 22 <213> ORGANISM: Homo sapiens  
 24 <220> FEATURE:  
 25 <221> NAME/KEY: CDS  
 26 <222> LOCATION: (1)..(1140)  
 28 <400> SEQUENCE: 1  
 29 atg gcc tcc ctt gct gca gca aat gca gag ttt tgc ttc aac ctg ttc 48  
 30 Met Ala Ser Leu Ala Ala Ala Asn Ala Glu Phe Cys Phe Asn Leu Phe  
 31 1 5 10 15  
 33 aga gag atg gat gac aat caa gga aat gga aat gtg ttc ttt tcc tct 96  
 34 Arg Glu Met Asp Asp Asn Gln Gly Asn Gly Asn Val Phe Phe Ser Ser  
 35 20 25 30  
 37 ctg agc ctc ttc gct gcc ctg gcc ctg gtc cgc ttg ggc gct caa gat 144  
 38 Leu Ser Leu Phe Ala Ala Leu Ala Leu Val Arg Leu Gly Ala Gln Asp  
 39 35 40 45  
 41 gac tcc ctc tct cag att gat aag ttg ctt cat gtt aac act gcc tca 192  
 42 Asp Ser Leu Ser Gln Ile Asp Lys Leu Leu His Val Asn Thr Ala Ser  
 43 50 55 60  
 45 gga tat gga aac tct tct aat agt cag tca ggg ctc cag tct caa ctg 240  
 46 Gly Tyr Gly Asn Ser Ser Asn Ser Gln Ser Gly Leu Gln Ser Gln Leu  
 47 65 70 75 80  
 49 aaa aga gtt ttt tct gat ata aat gca tcc cac aag gat tat gat ctc 288  
 50 Lys Arg Val Phe Ser Asp Ile Asn Ala Ser His Lys Asp Tyr Asp Leu  
 51 85 90 95  
 53 agc att gtg aat ggg ctt ttt gct gaa aaa gtg tat ggc ttt cat aag 336  
 54 Ser Ile Val Asn Gly Leu Phe Ala Glu Lys Val Tyr Gly Phe His Lys  
 55 100 105 110  
 57 gac tac att gag tgt gcc gaa aaa tta tac gat gcc aaa gtg gag cga 384  
 58 Asp Tyr Ile Glu Cys Ala Glu Lys Leu Tyr Asp Ala Lys Val Glu Arg  
 59 115 120 125  
 61 gtt gac ttt acg aat cat tta gaa gac act aga cgt aat att aat aag 432  
 62 Val Asp Phe Thr Asn His Leu Glu Asp Thr Arg Arg Asn Ile Asn Lys

## RAW SEQUENCE LISTING

DATE: 10/18/2002

PATENT APPLICATION: US/09/508,997A

TIME: 10:41:41

Input Set : D:\seqlist.txt

Output Set: N:\CRF4\10182002\I508997A.raw

```

63      130      135      140
65 tgg gtt gaa aat gaa aca cat ggc aaa atc aag aac gtg att ggt gaa 480
66 Trp Val Glu Asn Glu Thr His Gly Lys Ile Lys Asn Val Ile Gly Glu
67 145      150      155      160
69 ggt ggc ata agc tca tct gct gta atg gtg ctg gtg aat gct gtg tac 528
70 Gly Gly Ile Ser Ser Ala Val Met Val Leu Val Asn Ala Val Tyr
71      165      170      175
73 ttc aaa ggc aag tgg caa tca gcc ttc acc aag agc gaa acc ata aat 576
74 Phe Lys Gly Lys Trp Gln Ser Ala Phe Thr Lys Ser Glu Thr Ile Asn
75      180      185      190
77 tgc cat ttc aaa tct ccc aag tgc tct ggg aag gca gtc gcc atg atg 624
78 Cys His Phe Lys Ser Pro Lys Cys Ser Gly Lys Ala Val Ala Met Met
79      195      200      205
81 cat cag gaa cgg aag ttc aat ttg tct gtt att gag gac cca tca atg 672
82 His Gln Glu Arg Lys Phe Asn Leu Ser Val Ile Glu Asp Pro Ser Met
83      210      215      220
85 aag att ctt gag ctc aga tac aat ggt ggc ata aac atg tac gtt ctg 720
86 Lys Ile Leu Glu Leu Arg Tyr Asn Gly Gly Ile Asn Met Tyr Val Leu
87 225      230      235      240
89 ctg cct gag aat gac ctc tct gaa att gaa aac aaa ctg acc ttt cag 768
90 Leu Pro Glu Asn Asp Leu Ser Glu Ile Glu Asn Lys Leu Thr Phe Gln
91      245      250      255
93 aat cta atg gaa tgg acc aat cca agg cga atg acc tct aag tat gtt 816
94 Asn Leu Met Glu Trp Thr Asn Pro Arg Arg Met Thr Ser Lys Tyr Val
95      260      265      270
97 gag gta ttt ttt cct cag ttc aag ata gag aag aat tat gaa atg aaa 864
98 Glu Val Phe Phe Pro Gln Phe Lys Ile Glu Lys Asn Tyr Glu Met Lys
99      275      280      285
101 caa tat ttg aga gcc cta ggg ctg aaa gat atc ttt gat gaa tcc aaa 912
102 Gln Tyr Leu Arg Ala Leu Gly Leu Lys Asp Ile Phe Asp Glu Ser Lys
103      290      295      300
105 gca gat ctc tct ggg att gct tcg ggg ggt cgt ctg tat ata tca agg 960
106 Ala Asp Leu Ser Gly Ile Ala Ser Gly Gly Arg Leu Tyr Ile Ser Arg
107 305      310      315      320
109 atg atg cac aaa tct tac ata gag gtc act gag gag ggc acc gag gct 1008
110 Met Met His Lys Ser Tyr Ile Glu Val Thr Glu Glu Gly Thr Glu Ala
111      325      330      335
113 act gct gcc aca gga agt aat att gta gaa aag caa ctc cct cag tcc 1056
114 Thr Ala Ala Thr Gly Ser Asn Ile Val Glu Lys Gln Leu Pro Gln Ser
115      340      345      350
117 acg ctg ttt aga gct gac cac cca ttc cta ttt gtt atc agg aag gat 1104
118 Thr Leu Phe Arg Ala Asp His Pro Phe Leu Phe Val Ile Arg Lys Asp
119      355      360      365
121 gac atc atc tta ttc agt ggc aaa gtt tct tgc cct tga 1143
122 Asp Ile Ile Leu Phe Ser Gly Lys Val Ser Cys Pro
123      370      375      380
126 <210> SEQ ID NO: 2
127 <211> LENGTH: 380
128 <212> TYPE: PRT

```

## RAW SEQUENCE LISTING

DATE: 10/18/2002

PATENT APPLICATION: US/09/508,997A

TIME: 10:41:41

Input Set : D:\seqlist.txt

Output Set: N:\CRF4\10182002\I508997A.raw

```

129 <213> ORGANISM: Homo sapiens
131 <400> SEQUENCE: 2
132 Met Ala Ser Leu Ala Ala Ala Asn Ala Glu Phe Cys Phe Asn Leu Phe
133   1           5           10           15
135 Arg Glu Met Asp Asp Asn Gln Gly Asn Gly Asn Val Phe Phe Ser Ser
136           20           25           30
138 Leu Ser Leu Phe Ala Ala Leu Ala Leu Val Arg Leu Gly Ala Gln Asp
139           35           40           45
141 Asp Ser Leu Ser Gln Ile Asp Lys Leu Leu His Val Asn Thr Ala Ser
142           50           55           60
144 Gly Tyr Gly Asn Ser Ser Asn Ser Gln Ser Gly Leu Gln Ser Gln Leu
145   65           70           75           80
147 Lys Arg Val Phe Ser Asp Ile Asn Ala Ser His Lys Asp Tyr Asp Leu
148           85           90           95
150 Ser Ile Val Asn Gly Leu Phe Ala Glu Lys Val Tyr Gly Phe His Lys
151           100          105          110
153 Asp Tyr Ile Glu Cys Ala Glu Lys Leu Tyr Asp Ala Lys Val Glu Arg
154           115          120          125
156 Val Asp Phe Thr Asn His Leu Glu Asp Thr Arg Arg Asn Ile Asn Lys
157           130          135          140
159 Trp Val Glu Asn Glu Thr His Gly Lys Ile Lys Asn Val Ile Gly Glu
160  145          150          155          160
162 Gly Gly Ile Ser Ser Ser Ala Val Met Val Leu Val Asn Ala Val Tyr
163           165          170          175
165 Phe Lys Gly Lys Trp Gln Ser Ala Phe Thr Lys Ser Glu Thr Ile Asn
166           180          185          190
168 Cys His Phe Lys Ser Pro Lys Cys Ser Gly Lys Ala Val Ala Met Met
169           195          200          205
171 His Gln Glu Arg Lys Phe Asn Leu Ser Val Ile Glu Asp Pro Ser Met
172           210          215          220
174 Lys Ile Leu Glu Leu Arg Tyr Asn Gly Gly Ile Asn Met Tyr Val Leu
175  225          230          235          240
177 Leu Pro Glu Asn Asp Leu Ser Glu Ile Glu Asn Lys Leu Thr Phe Gln
178           245          250          255
180 Asn Leu Met Glu Trp Thr Asn Pro Arg Arg Met Thr Ser Lys Tyr Val
181           260          265          270
183 Glu Val Phe Phe Pro Gln Phe Lys Ile Glu Lys Asn Tyr Glu Met Lys
184           275          280          285
186 Gln Tyr Leu Arg Ala Leu Gly Leu Lys Asp Ile Phe Asp Glu Ser Lys
187           290          295          300
189 Ala Asp Leu Ser Gly Ile Ala Ser Gly Gly Arg Leu Tyr Ile Ser Arg
190  305          310          315          320
192 Met Met His Lys Ser Tyr Ile Glu Val Thr Glu Glu Gly Thr Glu Ala
193           325          330          335
195 Thr Ala Ala Thr Gly Ser Asn Ile Val Glu Lys Gln Leu Pro Gln Ser
196           340          345          350
198 Thr Leu Phe Arg Ala Asp His Pro Phe Leu Phe Val Ile Arg Lys Asp
199           355          360          365
201 Asp Ile Ile Leu Phe Ser Gly Lys Val Ser Cys Pro

```

## RAW SEQUENCE LISTING

DATE: 10/18/2002

PATENT APPLICATION: US/09/508,997A

TIME: 10:41:41

Input Set : D:\seqlist.txt

Output Set: N:\CRF4\10182002\I508997A.raw

```

202      370      375      380
205 <210> SEQ ID NO: 3
206 <211> LENGTH: 1229
207 <212> TYPE: DNA
208 <213> ORGANISM: Rattus rattus
210 <220> FEATURE:
211 <221> NAME/KEY: CDS
212 <222> LOCATION: (8)..(1147)
214 <221> NAME/KEY: unsure
215 <222> LOCATION: (158)..(160), (287)..(289)
216 <223> OTHER INFORMATION: n is unknown.
W--> 218 <400> 3
219 ttccaat atg gcc tcc ctt gct gca gca aat gca gaa ttt ggc ttc gac 49
220      Met Ala Ser Leu Ala Ala Ala Asn Ala Glu Phe Gly Phe Asp
221      1      5      10
223 tta ttc aga gag atg gat agt agt caa gga aac gga aat gta ttc ttc 97
224 Leu Phe Arg Glu Met Asp Ser Ser Gln Gly Asn Gly Asn Val Phe Phe
225 15      20      25      30
227 tct tcc ctg agc atc ttc act gcc ctg agc cta atc cgt ttg ggt gct 145
228 Ser Ser Leu Ser Ile Phe Thr Ala Leu Ser Leu Ile Arg Leu Gly Ala
229      35      40      45
W--> 231 cga ggt gac tgt nnn cgt cag att gac aag gcc ctg cac ttt atc tcc 193
W--> 232 Arg Gly Asp Cys Xaa Arg Gln Ile Asp Lys Ala Leu His Phe Ile Ser
233      50      55      60
235 cca tca aga caa ggg aat tca tcg aac agt cag cta gga ctg caa tat 241
236 Pro Ser Arg Gln Gly Asn Ser Ser Asn Ser Gln Leu Gly Leu Gln Tyr
237      65      70      75
W--> 239 caa ttg aaa aga gtt ctt gct gac ata aac tca tct cat aag gat nnn 289
W--> 240 Gln Leu Lys Arg Val Leu Ala Asp Ile Asn Ser Ser His Lys Asp Xaa
241      80      85      90
243 aaa ctc agc att gcc aat gga gtt ttt gca gag aaa gta ttt gat ttt 337
244 Lys Leu Ser Ile Ala Asn Gly Val Phe Ala Glu Lys Val Phe Asp Phe
245 95      100      105      110
247 cat aag agc tat atg gag tgt gct gaa aac tta tac aat gct aaa gtg 385
248 His Lys Ser Tyr Met Glu Cys Ala Glu Asn Leu Tyr Asn Ala Lys Val
249      115      120      125
251 gaa aga gtt gat ttt aca aat gat ata caa gaa acc aga ttt aaa att 433
252 Glu Arg Val Asp Phe Thr Asn Asp Ile Gln Glu Thr Arg Phe Lys Ile
253      130      135      140
255 aat aaa tgg att gaa aat gaa aca cat ggc aaa atc aag aag gtg ttg 481
256 Asn Lys Trp Ile Glu Asn Glu Thr His Gly Lys Ile Lys Lys Val Leu
257      145      150      155
259 ggg gac agc agc ctc agc tca tca gct gtc atg gtg cta gtg aat gct 529
260 Gly Asp Ser Ser Leu Ser Ser Ser Ala Val Met Val Leu Val Asn Ala
261      160      165      170
263 gtt tac ttc aaa ggc aag tgg aaa tcg gcc ttc acc aag agt gat acc 577
264 Val Tyr Phe Lys Gly Lys Trp Lys Ser Ala Phe Thr Lys Ser Asp Thr
265 175      180      185      190
267 ctc agt tgc cat ttc agg tct ccc agc ggt cct gga aaa gca gtt aat 625

```

## RAW SEQUENCE LISTING

DATE: 10/18/2002

PATENT APPLICATION: US/09/508,997A

TIME: 10:41:41

Input Set : D:\seqlist.txt

Output Set: N:\CRF4\10182002\I508997A.raw

```

268 Leu Ser Cys His Phe Arg Ser Pro Ser Gly Pro Gly Lys Ala Val Asn
269          195          200          205
271 atg atg cat caa gaa cgg agg ttc aat ttg tct acc att cag gag cca 673
272 Met Met His Gln Glu Arg Arg Phe Asn Leu Ser Thr Ile Gln Glu Pro
273          210          215          220
275 cca atg cag att ctt gag cta caa tat cat ggt ggc ata agc atg tac 721
276 Pro Met Gln Ile Leu Glu Leu Gln Tyr His Gly Gly Ile Ser Met Tyr
277          225          230          235
279 atc atg ttg ccc gag gat gac cta tcc gaa att gaa agc aag ctg agt 769
280 Ile Met Leu Pro Glu Asp Asp Leu Ser Glu Ile Glu Ser Lys Leu Ser
281          240          245          250
283 ttc cag aat cta atg gac tgg aca aat agc agg aag atg aaa tct cag 817
284 Phe Gln Asn Leu Met Asp Trp Thr Asn Ser Arg Lys Met Lys Ser Gln
285 255          260          265          270
287 tat gtg aat gtg ttt ctc ccc cag ttc aag ata gag aaa gat tat gaa 865
288 Tyr Val Asn Val Phe Leu Pro Gln Phe Lys Ile Glu Lys Asp Tyr Glu
289          275          280          285
291 atg agg agc cac ttg aaa tct gta ggc ttg gaa gac atc ttt gtt gag 913
292 Met Arg Ser His Leu Lys Ser Val Gly Leu Glu Asp Ile Phe Val Glu
293          290          295          300
295 tcc agg gct gat ctg tct gga att gcc tct gga ggt cgt ctc tat gta 961
296 Ser Arg Ala Asp Leu Ser Gly Ile Ala Ser Gly Gly Arg Leu Tyr Val
297          305          310          315
299 tca aag cta atg cac aag tcc ctc ata gag gtc tca gaa gaa ggc acc 1009
300 Ser Lys Leu Met His Lys Ser Leu Ile Glu Val Ser Glu Glu Gly Thr
301          320          325          330
303 gag gca act gct gcc aca gaa agt aac atc gtt gaa aag cta ctt cct 1057
304 Glu Ala Thr Ala Ala Thr Glu Ser Asn Ile Val Glu Lys Leu Leu Pro
305 335          340          345          350
307 gaa tcc acg gtg ttc aga gct gac cgc ccc ttt ctg ttt gtc att agg 1105
308 Glu Ser Thr Val Phe Arg Ala Asp Arg Pro Phe Leu Phe Val Ile Arg
309          355          360          365
311 aag aat ggc atc atc tta ttt act ggc aaa gtc tcg tgt cct 1147
312 Lys Asn Gly Ile Ile Leu Phe Thr Gly Lys Val Ser Cys Pro
313          370          375          380
315 tgaaattcta tttggttttc catacactaa caggcatgaa gaaacatcat aagtgaatag 1207
317 aattgtaatt ggaagtacat gg 1229
320 <210> SEQ ID NO: 4
321 <211> LENGTH: 380
322 <212> TYPE: PRT
323 <213> ORGANISM: Rattus rattus
325 <220> FEATURE:
326 <221> NAME/KEY: UNSURE
327 <222> LOCATION: 51, 94
328 <223> OTHER INFORMATION: Xaa is unknown.
330 <400> SEQUENCE: 4
331 Met Ala Ser Leu Ala Ala Ala Asn Ala Glu Phe Gly Phe Asp Leu Phe
332 1 5 10 15
334 Arg Glu Met Asp Ser Ser Gln Gly Asn Gly Asn Val Phe Phe Ser Ser

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/09/508,997A

DATE: 10/18/2002  
TIME: 10:41:42

Input Set : D:\seqlist.txt  
Output Set: N:\CRF4\10182002\I508997A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:3; N Pos. 158,159,160,287,288,289

Seq#:3; Xaa Pos. 51,94

Seq#:4; Xaa Pos. 51,94

Seq#:24; N Pos. 26,29

Seq#:25; N Pos. 3,9,15

Seq#:26; N Pos. 6,9,12,15,18,21